Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Chemical Biology

Protein TAILS: when termini tell tales of proteolysis and function

Philipp F Lange[1,2,3] and Christopher M Overall[1,2,3]

Among the hundreds of posttranslational modifications, limited proteolysis, also known as processing, is special: It is irreversible, near ubiquitous, and by trimming peptide chains from their ends or cutting proteins into two, proteolysis forms shorter chains displaying new termini. The unique chemistry and location of α-amino-termini and carboxyl-termini in a protein engender special chemical and physical properties to a protein. Hence, modification of protein termini is often associated with new biological activities of a protein. We highlight recent proteomic developments enabling high throughput identification of protein termini. This has revolutionized degradomics and protein characterization by mapping the specificity of terminal modifications and of proteases, and has been used to directly identify new protease substrates and molecular pathways altered by proteolysis.

Addresses

[1] University of British Columbia, Department of Oral Biological and Medical Sciences, 4.401 Life Sciences Institute, 2350 Health Sciences Mall, Vancouver, BC, Canada V6T 1Z3

[2] University of British Columbia, Department of Biochemistry and Molecular Biology, 4.401 Life Sciences Institute, 2350 Health Sciences Mall, Vancouver, BC, Canada V6T 1Z3

[3] University of British Columbia, Centre for Blood Research, 4.401 Life Sciences Institute, 2350 Health Sciences Mall, Vancouver, BC, Canada V6T 1Z3

Corresponding author: Overall, Christopher M (chris.overall@ubc.ca)

Current Opinion in Chemical Biology 2013, **17**:73–82

This review comes from a themed issue on **Omics**

Edited by **Matthew Bogyo** and **Pauline M Rudd**

For a complete overview see the Issue and the Editorial

Available online 6th January 2013

1367-5931

http://dx.doi.org/10.1016/j.cbpa.2012.11.025

Introduction

Proteomics has made astonishing advances in all areas including peptide enrichment, fractionation, mass spectrometry and data analysis — many of which are reviewed in this issue. With mass spectrometry based proteomics it is now possible to identify more than 10,000 proteins from human cells [1,2]. Recent advances in the field of protein post-translational modifications (PTMs) have uncovered their widespread occurrence and physiological relevance. However, for comprehensive analysis of PTMs specific peptide enrichment approaches and dedicated analyses are required, without which PTMs are usually undersampled and overlooked, respectively. In the absence of functional annotation of proteins from PTMs many key functions of bioactive proteins will be opaque and hence hypotheses based on traditional shotgun analyses, may be misleading or even worse, totally wrong.

PTM of proteins constitutes a highly diverse and dynamic regulatory layer affecting all aspects of a protein from protein folding, localization, interaction and bioactivity to its stability and ultimately degradation. Therefore, each distinctly modified version of a protein, also called a protein species, and not just the initial translated version, needs to be considered as the functional units comprising the proteome [3]. The diversity of reversible and irreversible modifications as well as the extensive modification machinery [4] and the possibility of combinatorial effects dramatically increase proteome complexity by several orders. Organisms as different as worm, fly and man have comparable sized genomes yet show a great discrepancy in phenotypic complexity. While splicing introduces bulk complexity it might well be that the diversity created by pinpoint posttranslational modifications accounts for the observed phenotypic differences. Hence, advanced proteomics has potential to explain phenotypes where conventional genomics fall short — but it is not easy.

Every modification adds to the functional diversity of the proteome by reversibly or irreversibly converting one protein species into another that potentially is a functionally distinct species. In this regard, limited proteolysis is special as it has the unique ability to irreversibly convert one into two distinct protein species while at the same time generating new protein termini serving as attachment sites for even further PTM.

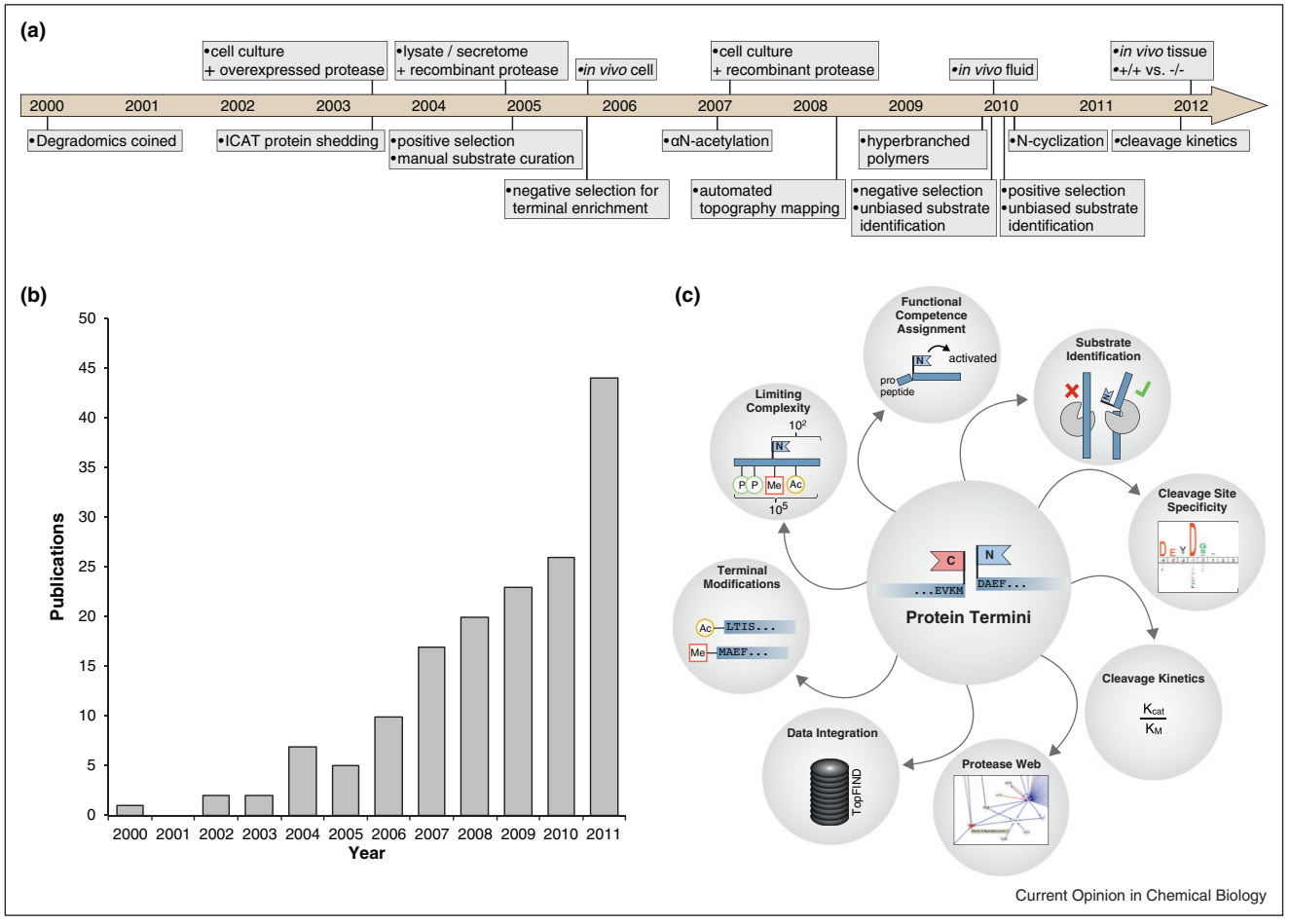
Second only to ubiquitin ligases in number, proteases and their inhibitors constitute a large enzyme family with 567 members in humans. In what has been termed the degradome, the assembly of all elements involved in proteolysis — proteases, inhibitors and the processed substrates — can now be specifically studied in high throughput investigations termed degradomics [5••]. Proteases modify their substrates by hydrolysis of scissile bonds releasing two peptide chains with the two amino acids adjacent to the cleaved bond now becoming carboxy-terminal or amino-terminal residues. Unlike most PTM attachment sites, the hydrolyzed peptide bond is not amenable for direct assessment. For limited proteolysis, termed processing, the site of modification is therefore determined by identification of the 'neo' termini of the products. Consequently the degradome and terminome are mutually dependent, with the identification of termini adding considerable functional annotation to the proteome.

Enriching for post-translational modifications

PTMs often occur at low stoichiometry and thus efficient enrichment techniques are key for their successful and comprehensive identification. In general different chemical affinities between the modified and unmodified species are utilized for differential binding to a resin or chromatographic media yielding positive or negative selectivity and enrichment. All approaches share the common hurdle of unspecific carryover and loss following binding to surfaces. A great advance for the enrichment of peptides bearing PTMs is the replacement of resins by soluble hyper branched polyglycerol polymers leading to massively decreased nonspecific binding while increasing binding capacity [6•].

Upon successful peptide enrichment mass spectrometry is used for peptide and PTM identification. Unlike identification of the entire protein by multiple peptides in one shotgun experiment, identification of a specific modification and often the protein bearing the PTM, is based on the observation of one single peptide only. For proteins having two or more such modifications, protein identification can often be made by two or more different and unique peptides. However, for single peptides bearing a PTM, such as phosphopeptides, unambiguous protein identification is problematic. For the identification of protein termini we and others introduced high confidence protein identification from single peptide identifications based on multiple peptide variants [7,8]. In the past ten years since its introduction [5] degradomics and its subfield, terminomics, have developed from a small field covered by only a few publications a year to a vibrant community publishing over 40 papers in 2011 (Figure 1). For in depth comparison of available mass spectrometry based methods for the proteome-wide analysis of limited proteolysis and their subsequent modification we refer to a recent review by Huesgen and Overall [9••] and by the accompanying paper in this issue from the Gaevert laboratory [10].

Figure 1



Proteome-wide identification of protein termini over the decade: **(a)** timeline of key steps in degradomics and terminomics since its introduction. The top stream covers the type and complexity of the analytes, the bottom shows key advances. **(b)** Count of publications covering degradomics and terminomics over the last 10 years. **(c)** Summary of areas driven by advances in protein terminus identification covered in this review.

Characterization of the proteolytic machinery

Since the function of a protease is inherently linked to the effect of proteolysis on its substrates, and since more than half of all proteases have no annotated substrates in MEROPS, the protease database (http://merops.sanger.ac.uk), since 2000 a major focus has been in the identification of protease substrates [11]. These include matrix metalloproteinases (MMPs) 2, 9, 14, 25 [6•,12–16], cathepsins D and E [8,17] and caspases 2, 3, 7 [18], meprins, astacins, and the methionine aminopeptidase-2 [19]. *In vivo* the cleavage rate differs greatly between individual substrates by the same protease [20•]. The cleavage site specificity of proteases has been investigated in depth using standard and specifically tailored degradomics approaches using database searchable, proteome-derived peptide libraries in a procedure called PICS [21]. This leads to the detailed characterization of both the prime and nonprime specificity of, for example, MMPs [21], cathepsins [22], astacins [23], caspases [18,21] and methionine aminopeptidase-1 [24,25••]. However, specific analysis of cleavage sites in denatured peptides does not identify native substrates. For this task, COFRADIC [26] and TAILS [6•] are highly successful negative selection approaches that also provide information on the nature of posttranslational modifications of termini.

Of particular importance for the characterization of any enzyme is the assessment of its kinetic properties *in vivo*. Employing identification of protease derived termini followed by time resolved quantification by single reaction monitoring (SRM) Agard and colleagues monitored the cleavage kinetics of caspases in lysates and living cells [27••]. Compounding the problems in analysis is the fact that proteases do not act independently, but are interconnected in the protease web [28]. Comprehensive proteome-wide analysis of global proteolysis by terminomics in complex mammalian tissues, comparing protease knockout mice with wild types, is now enabled for the first time for the *in vivo* investigation of such network effects [29••]. Hence, degradomics has advanced considerably from the first experimental paper in 2004 presenting ICAT labeled protein fragments shed from membrane proteins [12].

Protein termini expose new interfaces

Despite methodological advances, data without context is information, not knowledge. To combine the ever-growing body of information on protein termini and limited proteolysis, to discover network effects and integrate this with prior knowledge the 'Termini oriented protein function inferred database' (TopFIND — http://clipserve.clip.ubc.ca/topfind) [30] acts as central repository and information resource. Thereby, an evaluation of thirteen terminomics datasets from *Homo sapiens*, *Mus musculus*, and *Escherichia coli* shows that >30% of all N-termini and >10% of all C-termini originate from post-translational proteolytic processing other than classical protein maturation (removal of the initiator methionine, signal peptide and pro-peptide) [31•]. More recently, in skin 50% of the >2000 proteins identified had evidence of stable cleavage products *in vivo* [29••].

Terminal regions of a protein are often flexible, protruding and distinct from internal, continuous amino acid stretches and therefore frequently act as recognition sites for receptors and antibodies. Thus, by frequent formation of new N-termini or C-termini, limited proteolysis closes interfaces while opening up new ones that can be further altered by amino acid modifications ranging from post-translational acetylation to cyclization or palmitoylation.

While several hundred PTMs are listed in Unimod, which serves as the comprehensive reference database for protein modifications (http://www.unimod.org), certain modifications are specific to the free amino or carboxyl terminus and thus can only occur at one site each in a protein. Thus, terminomics has been extensively used to study co-translational and post-translational modification of terminal amino acids.

N-terminal modifications

In addition to co-translational acetylation Unimod and TopFIND report 11 amino-terminal PTMs including acetylation, mono-methylation, di-methylation and tri-methylation, formylation, carbamylation, succinylation, cyclization, propionylation, palmitoylation and myristoylation. Among these αN-acetylation and cyclization are the only two studied in depth at the mechanistic and proteome-wide level. αN-acetylation plays an important regulatory role in protein stability and protein turnover via the N-end rule [32••,33]. Initially, only co-translational αN-acetylation was recognized. However, post-translational αN-acetylation is now recognized as widespread PTM occurring *in vivo* [6•,25••,29••,34,35,36•,37]. Comprehensive understanding of αN-acetylation also enables the identification of alternate translational start sites utilizing the differential patterns displayed by co-translational and post-translational αN-acetylation [29••].

Chen *et al.* [38] for the first time described a physiological function for terminal methylation. The binding efficiency of regulator of chromatin condensation 1 (RCC1) to H2A and/or H2B depends on its terminal methylation with defective methylation leading to spindle-pole defects. Interestingly recent experiments suggest functional interplay or competition between αN-acetylation and αN-methylation [39] and probably also αN-propionylation, which early terminomics studies identified as occurring *in vivo* [34].

Cyclization of a terminal glutaminyl or glutamate residue forming N-pyroglutamate, a process initially believed to occur spontaneously but now recognized to be catalyzed by two glutaminylcyclases [40], attracts interest in

Alzheimer's research following the identification of a toxic pyroglutamate modified APP Aβ species (see below). *In vivo* sequence specificity for N-terminal cyclization has now been recently determined by TAILS, which enriches for all blocked termini, regardless of the modification [29••]. Finally, attachment of fatty acid or prenyl moieties is not strictly limited to the terminal amino acid, but N-myristoylation and N-palmitoylation are known to mediate signaling and trafficking [41,42], making their location at a terminus special, as these sites and cell attachment can be lost upon cleavage. While this group of PTMs has been extensively studied by classical biochemical and cell biological approaches its proteome wide relevance remains to be shown.

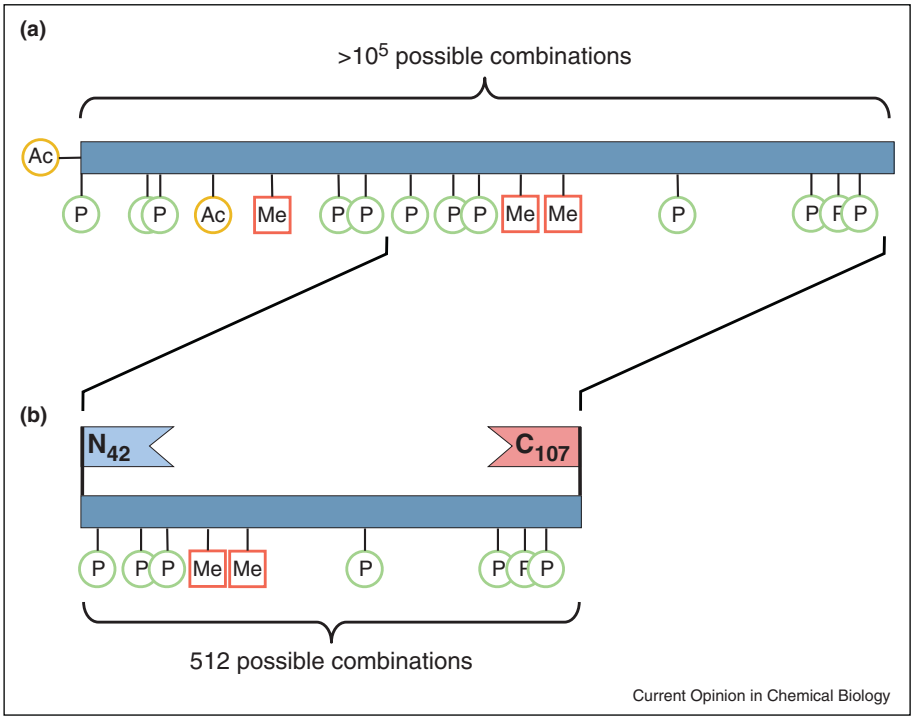
C-terminal modifications

The C terminus of proteins is inherently less reactive than the N terminus. While this may lead to less extensive modification in nature, this too is the very reason for the lack of C-terminal sequencing ability and hence recognition of C terminal modifications, until recently. Even with the recent advent of high throughput identification of C-termini [34,43,44] we still have very limited understanding of modifications affecting the carboxy-terminus.

C-methyl-esterification is the most frequently annotated modification, but with only 17 proteins in human, 6 in mouse and 7 in yeast reported by TopFIND, yet the C-termini remain underexplored. Examples include the methylation of the C-terminal leucine residue in the serine-threonine phosphatase 2A catalytic subunit ($PP2A_c$), which is required for the interaction with its regulatory Bα subunit [45].

C-terminal isoprenylation, cholesterol-esterification and addition of GPI anchors are involved in membrane targeting and trafficking but most of these were studied by classical biochemical analyses over the past 20 years [46•]. We suggest that the limited number of described C-terminal PTMs does not reflect reality, but rather is due to the lack of appropriate technologies for the in depth analysis of C-termini and their modifications until recently. Given the high number of carboxypeptidases greatly exceeding what can possibly be needed for mere degradation, the identification of C-terminal processing

Figure 2



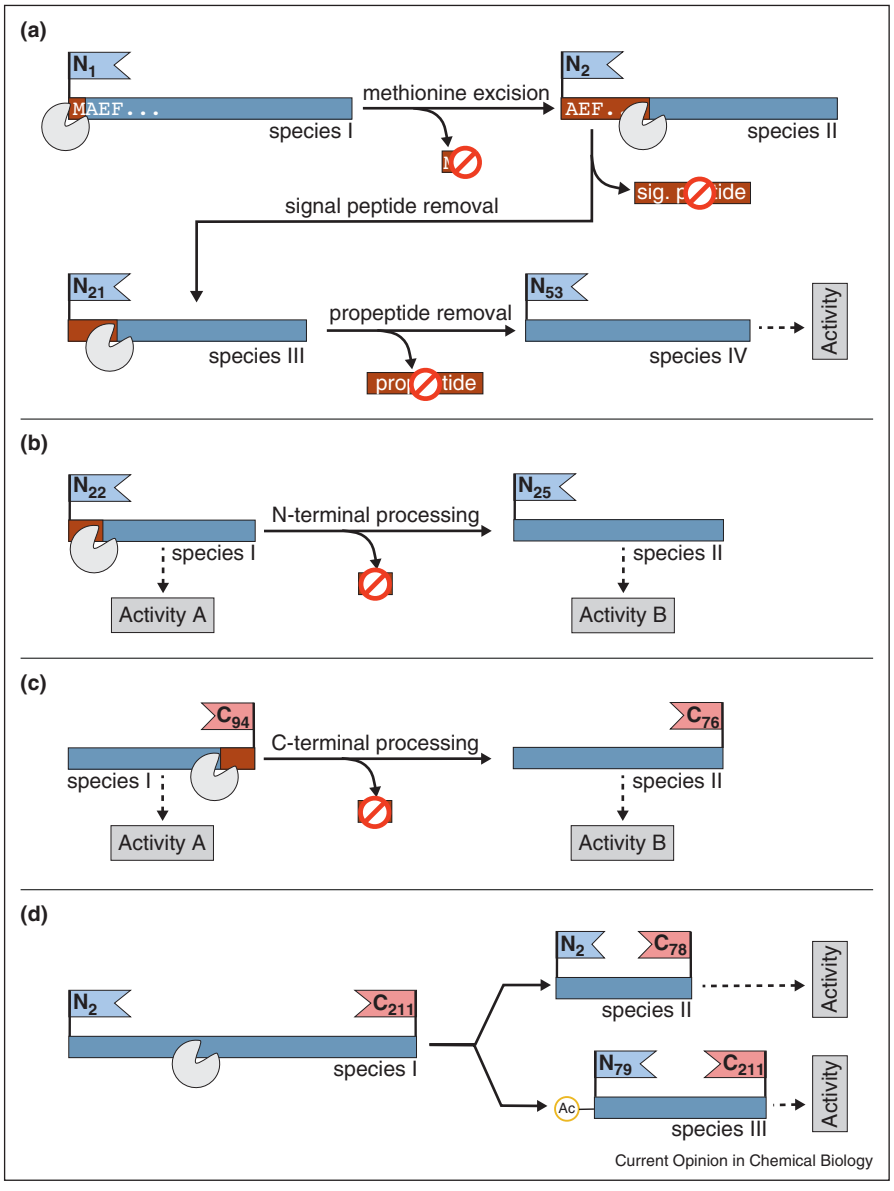
Using knowledge of protein termini to set boundaries and limit the complexity of the proteome introduced by combinatorial use of PTMs: **(a)** schematic depiction of HMGA1 and its possible attachment sites for post-translation modification by acetylation (Ac, yellow circle), phosphorylation (P, green circle), and mono-methylation or di-methylation (Me, red square) as annotated by neXtProt. Combinatorial use of all modification sites allows for the formation of $>10^5$ protein species. **(b)** HMGA1 after applying a boundary based on the identified protein terminus at position 42 (http://clipserve.clip.ubc.ca/topfind/proteins/P17096). Only 9 of the 17 possible PTM sites are located within these limits so reducing the possible number of protein species to 512 and narrowing the search space for de-convolution by 200 fold (blue flag, amino terminus at position 42 of the gene-encoded sequence; red flag, carboxy-terminus at position 107).

[43] and the physiological importance of the few modifications already known, in depth investigation of C-terminal modifications promises great potential for exiting new mechanistic insights into protein function.

Knowledge of protein termini limits complexity

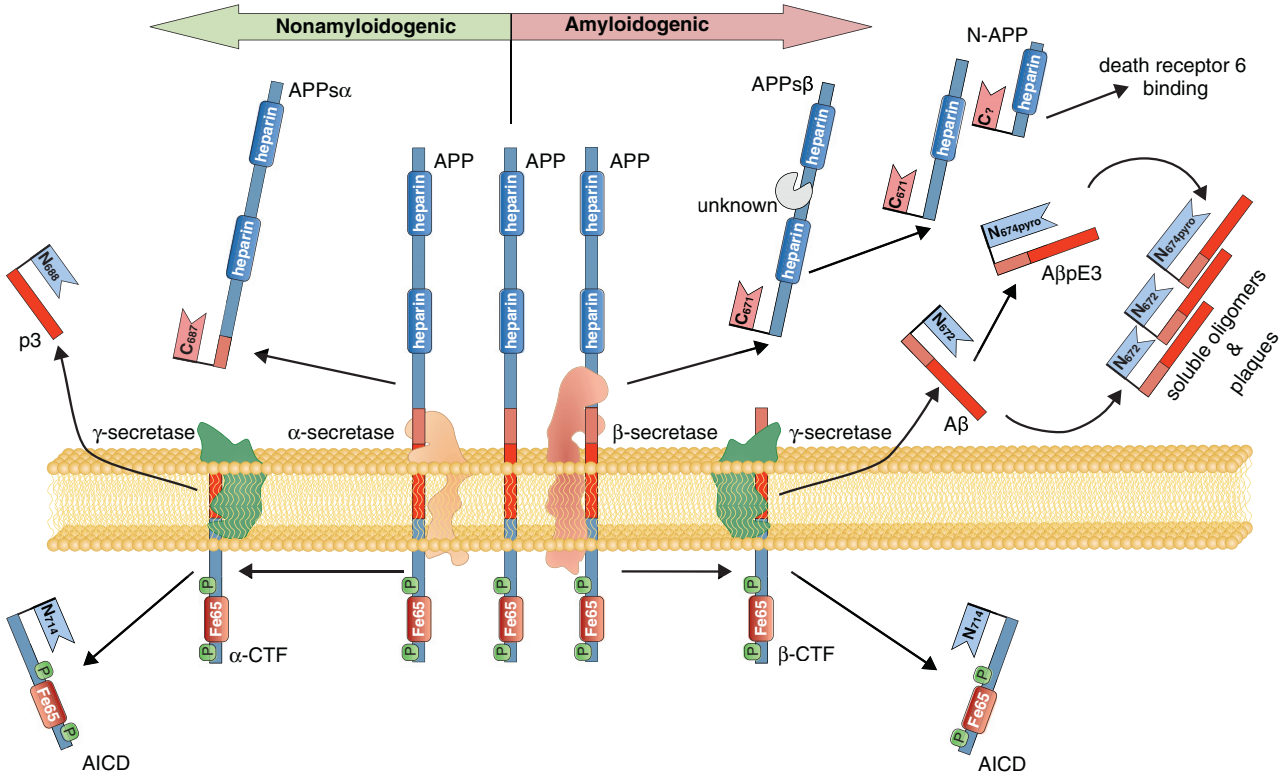
The notion that every PTM and combination thereof added to a protein needs to be considered as independent protein species led to the formulation of the histone code [47,48]. With several hundred distinct PTM sites described for histones alone this translates into mind-numbing complexity. While there is considerable debate about the *in vivo* relevance of PTM combinations [49•] recent work shows the *in vivo* presence, if not relevance, of multiple PTM combinations. Using top-down proteomics to map protein isoforms more than 100 protein species for the high mobility group (HMG) family of

Figure 3



Classes of protein modification by limited proteolysis: **(a)** sequential maturation. Single amino acids or peptides are excised from the main protein backbone by exo-peptidases or endo-peptidases, respectively. One of the fragments undergoes degradation without exerting intermediate activity resulting in sequential conversion of one protein species into another. **(b)** N-terminal trimming of one or several amino acids by exo-peptidases or endo-peptidases forming a new protein species with altered function. **(c)** C-terminal trimming of one or several amino acids by exo-peptidases or endo-peptidases forming a new protein species with altered function. **(d)** Protein partitioning. Endopeptidase activity forms two stable protein species (II, III), both able to exert potentially different activity. Flag symbols indicate the unique protein termini of the stable protein species (blue flag, N-terminus; red flag, C-terminus).

Figure 4



Name	Protein Species	Generated By	Functional Properties				Functional Competence	
			Present	Lost	Acquired	Localization	Known	Inferred
APP	N18-C770		Death Receptor 6 (DR6) binding site, heparin-binding domain, Fe65 binding site, X11 binding site, TM, sorting motifs, phosphorylation sites	signal peptide		membrane anchored, endosome, plasma membrane	intracellular domain mediates sorting, F-spondin binding to heparin-binding domain inhibits α- and β-secretase cleavage, can be processed into subsequent species	membrane anchored, potential for DR6 binding and signaling after direct cell-cell contact
AICD	N714-C770	γ-secretase,	Fe65 binding site, X11 binding site, sorting motifs, phosphorylation sites	DR6 binding site, heparin-binding domain, TM	soluble	cytosol, nucleus	degraded or translocated to nucleus for regulation of transcription	soluble, Fe65 / X11 binding and phosphorylation can affect stability and nuclear translocation
APPsα	N18-C687	α-secretase,	DR6 binding site, heparin-binding domain	TM, intracellular domains, sorting motifs, phosphorylation sites	soluble	extracellular space		not affected by intracellular APP sorting and phosphorylation, potential binding to DR6 and heparin-binding proteins
p3	N688-C713	α-secretase, γ-secretase		DR6 binding site, heparin-binding domain TM, intracellular domains	soluble	extracellular space		not affected by intracellular APP sorting and phosphorylation, no known functional domains present
APPsβ	N18-C671	β-secrease	DR6 binding site, heparin-binding domain	TM, intracellular domains, sorting motifs, phosphorylation sites	soluble	extracellular space	binding to DR6 activates caspase 6 and caspase 3 triggering axon pruning and neuron death	soluble, extracellular, not affected by intracellular APP sorting and phosphorylation, no known functional domains present
Aβ40	N672-C711	β-secretase, γ-secretase		DR6 binding site, heparin-binding domain, TM, intracellular domains	soluble	extracellular space	regulates synaptic scaling and synaptic vesicle release, potential to form soluble oligomers and plaques	soluble, extracellular, not affected by intracellular APP sorting and phosphorylation, no known functional domains present
Aβ42	N672-C713	α-secretase, γ-secretase		DR6 binding site, heparin-binding domain, TM, intracellular domains	soluble	extracellular space	regulates synaptic scaling and synaptic vesicle release, high potential to form soluble oligomers and plaques	no known functional domains soluble, extracellular, not affected by intracellular APP sorting and phosphorylation, no known functional domains present
AβpE3	N674pyro-C713	α-secretase, γ-secretase, Meprin β		DR6 binding site, heparin-binding domain, TM, intracellular domains	soluble, N-pyroglutamate, increased stability	extracellular space	increased stability, increased potential to form soluble oligomers and plaques	soluble, extracellular, not affected by intracellular APP sorting and phosphorylation, no known functional domains present

Current Opinion in Chemical Biology

Functional competence of a protein flagged by protein termini: simplified representation of processing of the amyloid precursor protein (APP) in the nonamyloidogenic (left, green) and amyloidogenic (right, red) pathways by α-secretases, β-secretases and γ-secretases. For full details see TopFIND (http://clipserve.clip.ubc.ca/topfind/proteins/P05067). In the non-pathologic nonamyloidogenic pathway APP is first cleaved by α-secretase near the plasma membrane. This releases the soluble APPsα species having a unique C_{687}-termninus and unknown function. Subsequent intramembrane cleavage by y-secretase generates the p3 peptide with a unique N_{688}-terminus and the APP intracellular domain, which is most likely degraded. The first processing step in the amyloidogenic pathway is mediated by β-secretase, BACE-1, in the endosome at a slightly different position to release APPsβ and Aβ species with the unique C_{671}-termini and N_{672}-termini, respectively. APPsβ undergoes further sequential maturation by an unknown protease at an undefined position generating the N-APP species, which binds to the death receptor 6 (DR6) on cells *in trans* triggering axon pruning

57 genes are known, including many containing multiple phosphorylations and methylations [50••]. Multiple modifications can cooperate by two fundamental principles. First, the total number of modifications can be critical to reach a certain threshold for a change in protein function. For example charge accumulation or masking alters the dipole moment of a molecule thereby attracting or repelling specific protein–protein interactions. Second, the exact combination of modifications can be required in order to reach a physiological outcome hence conveying true combinatorial specificity.

While distinct modification sites and identified species are now in the hundreds for histones and the HMG family, these numbers are dwarfed by the theoretical number of possible species formed by combinatorial use of PTM sites. Considering only HMGA1 and PTM sites annotated by neXtProt (http://nextprot.org) a total of $>10^5$ protein species could potentially exist (Figure 2a). In some cases an unmodified protein forms a reservoir of inactive protein awaiting activation by modification, in others the PTM switches activity of the protein from one type to another. Combined combinatorial considerations lead to many new species having graduations of activity.

While top-down proteomics provides direct identification of a protein species including all of its PTMs, assigning peptide identifications from shotgun analyses to specific protein species remains problematic. However, as exemplified in Figure 2b for a TopFIND analysis of HMGB1 (http://clipserve.clip.ubc.ca/topfind/proteins/P17096), knowledge of the terminal peptides of the species present in the sample provides boundaries drastically reducing the search space.

Protein termini are markers for functional competence

Modification of a protein by limited proteolysis can be divided into two general classes: first, sequential maturation and second, protein partitioning. During sequential maturation the removal of, for example, a propeptide that maintains enzyme latency, enables enzymatic activity of the major chain, but the propeptide, its task done, is most often then degraded (Figure 3a). Similarly, chemokine functions are frequently altered by truncation of few amino acids at their N-terminus or C-terminus (Figure 3b and c). CCL2 and CCL7, for example, become antagonists after N-terminal truncation [11]. In contrast, partitioning leads to the formation of two new protein species with usually unrelated properties thereby increasing the complexity of the proteome and potential for functional diversity (Figure 3d). HARP cleavage by MMP2 generates two bioactive species having opposing activity — the N-terminal species increases mitogenesis whereas the C-terminal species is antagonistic [13]. Irrespective of its mode of formation each new protein species is characterized by one 'neo' terminus. New functionality can be introduced by further modification of the new terminus including the recent recognition of post-translational acetylation [29••] thereby increasing the functional repertoire of the new protein species. However, as the species inherits only a subset of its progenitors features, such as active sites, binding regions and PTM sites, the potential functional complexity is limited.

In the following we use the amyloid beta A4 protein (APP) to illustrate how protein termini identified by terminomics can serve as markers for the functionality a protein species. We refer to this as the 'functional competence' of a protein species which can be obtained by 'positional cross correlation' of a species' termini with prior functional knowledge [31•].

APP is well known for its role in Alzheimer's disease [51]. APP is a single pass type-I transmembrane protein that undergoes a series of partitioning processing steps leading to multiple bioactive species (Figure 4). Comparing the normal nonamyloidogenic with disease causing amyloidogenic situations, the participation of different proteases in different subcellular compartments and facing changing physicochemical conditions translate to minute differences in species length and dramatic changes in systemic effect. Many more cleavages and species have been described for APP [52] (see TopFIND: http://clipserve.clip.ubc.ca/topfind/proteins/P05067#processing), for example of the N-APP species by meprin β [53]. However, even the simplified version depicted in Figure 4 highlights the complexity of the different functional and disease outcomes associated with different APP species. Clearly then, conventional shotgun proteomics cannot easily allow for distinguishing which APP species gave rise to a specific peptide and thus fails to capture this complexity and providing only incomplete information. This void can be filled by identification and quantification of the protein termini which allows not

(Figure 4 Legend Continued) and neuron death. Aβ in turn form oligomers or accumulate in sheets and is believed to be the major pathogenic component of the amyloidogenic pathway. The Aβ species can undergo further trimming at its amino-terminus followed by cyclization of the now terminal glutamate residue that is recently claimed to be the toxic species. This AβpE3 species is flagged by a unique $N_{674pyro}$ terminus and can initiate Aβ aggregation. Following β-secretase cleavage of APP, γ-secretase releases the AICD species to the cytoplasm, which is degraded or translocated to the nucleus regulating transcription. The key lists the major APP species, some are unambiguously identified by one of their termini (bold in legend, flag symbol in schema) associating the terminus with either the amyloidogenic (red background) or nonamyloidogenic (green background) pathway. On the basis of the presence or absence of functional groups and prior knowledge, the functional competence can be defined (last column) and associated with the flagged terminus. Flag symbols indicate the unique protein termini of the protein species (blue flag, N-terminus; red flag, C-terminus).

only for dissection of the different species present but also to infer their function and the proteolytic process by which they were generated. If the protease concerned is a drug target then specific monitoring of these terminal peptides forms an invaluable biomarker for drug efficacy and treatment progression.

Unfortunately knowledge of one terminus is not always enough to unambiguously identify a protein species. For example, to differentiate the pathophysiological differing species $A\beta_{40}{}^{N672\text{-}C711}$ and $A\beta_{42}{}^{N672\text{-}C713}$ additional knowledge of their C-terminus is required (Figure 4). Thus, in complex mixtures terminomics faces the same limitations as classical shotgun proteomics but while incredible advances have been made in top-down analyses [50••] they are still not readily available. So for now we can conclude that knowledge of the N-terminus and/or C-terminus reveals important information about a protein species including, first, the proteolytic processes leading to its formation; second, the protein features present and lost; third, its functional competence; fourth, and often its predicted stability and thus is essential for generating biologically relevant hypotheses on the protein's function *in vivo*.

Conclusions

The advent of proteomics for the enrichment and investigation of protein termini triggered a number of exiting findings and developments. First, when the investigation of limited proteolysis on a proteome-wide level became amenable this resulted in an explosion of newly identified protease substrates, an unexpected number of which are protease inhibitors or proteases themselves. This in turn enforced our understanding of proteolysis as a process occurring in a tightly interdependent network we have termed the protease web. With characterization of specificity and *in vivo* kinetics, great advances have also been made on the level of the individual enzyme *in vivo*.

One area that holds great potential, indeed well deserving of greater attention, is the global characterization of terminal modifications other than N-acetylation, in particular for carboxy-terminal modifications. These are particularly interesting and their unique properties — there is only one of each per protein chain and they only can come into existence after proteolytic processing — add a fascinating level of regulatory complexity.

A challenge of great enormity is to delve into the vast complexity engendered by PTMs. We have described how knowledge of protein termini will facilitate this by setting boundaries to the search space and acting as biomarkers defining the functional state of a protein. In the near future this will lead to exiting new biological insights into cellular and disease processes at a systems level and help close the gap between genotypes and phenotypes.

Acknowledgements

This work was supported by grants of the Canadian Institutes of Health Research; the Canadian Breast Cancer Research Alliance; the Canadian Breast Cancer Foundation; the Cancer Research Society; a Canada Research Chair to C.M.O., the Michael Smith Foundation for Health Research, the Breast Cancer Society of Canada, Alexander von Humboldt Foundation and the German Federal Ministry of Education and Research to P.F.L.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, Pääbo S, Mann M: **Deep proteome and transcriptome mapping of a human cancer cell line**. *Mol Syst Biol* 2011, **7**:1-8.

2. Beck M, Schmidt A, Malmstroem J, Claassen M, Ori A, Szymborska A, Herzog F, Rinner O, Ellenberg J, Aebersold R: **The quantitative proteome of a human cell line**. *Mol Syst Biol* 2011, **7**:1-8.

3. Jungblut PR, Holzhutter HG, Apweiler R, Schlüter H: **The speciation of the proteome**. *Chem Central J* 2008, **2**:16.

4. Walsh CT, Garneau-Tsodikova S, Gatto GJ: **Protein posttranslational modifications: the chemistry of proteome diversifications**. *Angew Chem Int Ed* 2005, **44**:7342-7372.

5. •• López-Otín C, Overall CM: **Protease degradomics: a new challenge for proteomics**. *Nat Rev Mol Cell Biol* 2002, **3**:509-519.
Introduction and description of methods to apply proteomics to proteases and substrate discovery.

6. • Kleifeld O, Doucet A, Keller UAD, Prudova A, Schilling O, Kainthan RK, Starr AE, Foster LJ, Kizhakkedathu JN, Overall CM: **Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products**. *Nat Biotechnol* 2010, **28**:281-288.
Introduction of novel polymers for proteomics and unbiased substrate discovery following negative selection of peptides including analysis of lysates, cells and biological fluids.

7. auf dem Keller U, Prudova A, Gioia M, Butler GS, Overall CM: **A statistics-based platform for quantitative N-terminome analysis and identification of protease cleavage products**. *Mol Cell Proteomics* 2010, **9**:912-927.

8. Impens F, Colaert N, Helsens K, Ghesquière B, Timmerman E, De Bock P-J, Chain BM, Vandekerckhove J, Gevaert K: **A quantitative proteomics design for systematic identification of protease cleavage events**. *Mol Cell Proteomics* 2010, **9**:2327-2333.

9. •• Huesgen PF, Overall CM: **N- and C-terminal degradomics: new approaches to reveal biological roles for plant proteases from substrate identification**. *Physiol Plant* 2012, **145**:5-17.
This review serves as an excellent introduction and reference to all major N-terminal and C-terminal enrichment methodologies.

10. Plasman K, Van Damme P, Gevaert K: **Contemporary positional proteomics strategies to study protein processing**. *Curr Opin Chem Biol*, http://dx.doi.org/10.1016/j.cbpa.2012.11.026.

11. McQuibban GA, Gong JH, Tam EM, McCulloch CA, Clark-Lewis I, Overall CM: **Inflammation dampened by gelatinase A cleavage of monocyte chemoattractant protein-3**. *Science* 2000, **289**:1202-1206.

12. Tam EM, Morrison CJ, Wu YI, Stack MS, Overall CM: **Membrane protease proteomics: isotope-coded affinity tag MS identification of undescribed MT1-matrix metalloproteinase substrates**. *Proc Natl Acad Sci U S A* 2004, **101**:6917-6922.

13. Dean RA, Butler GS, Hamma-Kourbali Y, Delbé J, Brigstock DR, Courty J, Overall CM: **Identification of candidate angiogenic inhibitors processed by matrix metalloproteinase 2 (MMP-2) in**

cell-based proteomic screens: disruption of vascular endothelial growth factor (VEGF)/heparin affin regulatory peptide (pleiotrophin) and VEGF/connective tissue growth factor angiogenic inhibitory complexes by MMP-2 proteolysis. *Mol Cell Biol* 2007, **27**:8454-8465.

14. Butler GS, Dean RA, Tam EM, Overall CM: **Pharmacoproteomics of a metalloproteinase hydroxamate inhibitor in breast cancer cells: dynamics of membrane type 1 matrix metalloproteinase-mediated membrane protein shedding**. *Mol Cell Biol* 2008, **28**:4896-4914.

15. Prudova A, auf dem Keller U, Butler GS, Overall CM: **Multiplex N-terminome analysis of MMP-2 and MMP-9 substrate degradomes by iTRAQ-TAILS quantitative proteomics**. *Mol Cell Proteomics* 2010, **9**:894-911.

16. Starr AE, Dufour A, Maier J, Overall CM: **Biochemical analysis of matrix metalloproteinase activation of chemokines CCL15 and CCL23 and increased glycosaminoglycan binding of CCL16**. *J Biol Chem* 2012, **287**:5848-5860.

17. Laurent-Matha V, Huesgen PF, Masson O, Derocq D, Prébois C, Gary-Bobo M, Lecaille F, Rebière B, Meurice G, Oréar C *et al.*: **Overall, and Emmanuelle Liaudet-Coopman Proteolysis of cystatin C by cathepsin D in the breast cancer microenvironment**. *FASEB J* 2012, **26**:5172-5181 http://dx.doi.org/10.1096/fj.12-205229.

18. Wejda M, Impens F, Takahashi N, Van Damme P, Gevaert K, Vandenabeele P: **Degradomics reveals that cleavage specificity profiles of caspase-2 and effector caspases are alike**. *J Biol Chem* 2012 http://dx.doi.org/10.1074/jbc.M112.384552.

19. Sundberg TB, Darricarrere N, Cirone P, Li X, McDonald L, Mei X, Westlake CJ, Slusarski DC, Beynon RJ, Crews CM: **Disruption of Wnt planar cell polarity signaling by aberrant accumulation of the MetAP-2 substrate Rab37**. *Chem Biol* 2011, **18**:1300-1311.

20. • Shimbo K, Hsu GW, Nguyen H, Mahrus S, Trinidad JC, Burlingame AL, Wells JA: **Quantitative profiling of caspase-cleaved substrates reveals different drug-induced and cell-type patterns in apoptosis**. *Proc Natl Acad Sci U S A* 2012 http://dx.doi.org/10.1073/pnas.1208616109.
By first applying N-terminal enrichment based discovery followed by SRM based targeted analysis the authors offer a glimpse on the proteolytic processes during apoptosis as they occur over time.

21. Schilling O, Overall CM: **Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites**. *Nat Biotechnol* 2008, **26**:685-694.

22. Biniossek ML, Nägler DK, Becker-Pauly C, Schilling O: **Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S**. *J Proteome Res* 2011, **10**:5363-5373.

23. Becker-Pauly C, Barré O, Schilling O, auf dem Keller U, Ohler A, Broder C, Schütte A, Kappelhoff R, Stöcker W, Overall CM: **Proteomic analyses reveal an acidic prime side specificity for the astacin metalloprotease family reflected by physiological substrates**. *Mol Cell Proteomics* 2011, **10** M111.009233.

24. Frottin F, Martinez A, Peynot P, Mitra S, Holz RC, Giglione C, Meinnel T: **The proteomics of N-terminal methionine cleavage**. *Mol Cell Proteomics* 2006, **5**:2336-2349.

25. •• Helbig AO, Gauci S, Raijmakers R, van Breukelen B, Slijper M, Mohammed S, Heck AJR: **Profiling of N-acetylated protein termini provides in-depth insights into the N-terminal nature of the proteome**. *Mol Cell Proteomics* 2010, **9**:928-939.
In a multi-species study the authors reveal that the impact of the penultimate position on cleavage efficiency is essentially conserved from *E. coli* to man and that sequence differences rather originate from genome differences.

26. Gevaert K, Goethals M, Martens L, Van Damme J, Staes A, Thomas GR, Vandekerckhove J: **Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides**. *Nat Biotechnol* 2003, **21**:566-569.

27. •• Agard NJ, Mahrus S, Trinidad JC, Lynn A, Burlingame AL, Wells JA: **Global kinetic analysis of proteolysis via quantitative targeted proteomics**. *Proc Natl Acad Sci U S A* 2012, **109**:1913-1918.
Employing first discovery by terminal enrichment followed by targeted monitoring of select termini over time the authors nicely demonstrate the power of terminomics and for the first time investigate kinetics of proteolysis on a broad scale *in vivo*.

28. Doucet A, Butler GS, Rodríguez D, Prudova A, Overall CM: **Metadegradomics: toward in vivo quantitative degradomics of proteolytic post-translational modifications of the cancer proteome**. *Mol Cell Proteomics* 2008, **7**:1925-1951.

29. •• Keller UAD, Prudova A, Eckhard U, Fingleton B, Overall CM: **Terminomics reveals MMP2 alters the protease web to increase vessel permeability and complement activity in skin inflammation**. Sci Signal 2012, in press
By four way cross comparison of healthy and inflamed skin from wild type and protease knock out mice the authors reveal detailed mechanistic insight on proteolysis in inflammation. The study also reports in depth analysis of terminal modifications, in particular pyro-glutamate formation, and identification of alternative translation initiation sites *in vivo*.

30. Lange PF, Huesgen PF, Overall CM: **TopFIND 2.0 — linking protein termini with proteolytic processing and modifications altering protein function**. *Nucleic Acids Res* 2011, **40**:D351-D361.

31. • Lange PF, Overall CM: **TopFIND, a knowledgebase linking protein termini with function**. *Nat Methods* 2011, **8**:703-704.
In this work the authors perform meta analysis of large scale terminomics studies from multiple laboratories reporting prevalence of limited proteolysis, *in vivo* applicability of the N-end rule and the need for detailed analysis of terminal modification other than acetylation. The authors also announce TopFIND, a central knowledgebase linking protein termini with proteolysis and function.

32. •• Varshavsky A: **The N-end rule pathway and regulation by proteolysis — Varshavsky — 2011 — Protein Science — Wiley Online Library**. *Protein Sci* 2011, **20**:1298-1345.
In this review Alexander Varshavsky provides a comprehensive overview on all aspects of the N-end rule pathway.

33. Sriram SM, Kim BY, Kwon YT: **The N-end rule pathway: emerging functions and molecular principles of substrate recognition**. *Nat Rev Mol Cell Biol* 2011, **12**:735-747.

34. Dormeyer W, Mohammed S, Breukelen BV, Krijgsveld J, Heck AJR: **Targeted analysis of protein termini**. *J Proteome Res* 2007, **6**:4634-4645.

35. Mischerikow N, Heck AJR: **Targeted large-scale analysis of protein acetylation**. *Proteomics* 2011, **11**:571-589.

36. • Van Damme P, Lasa M, Polevoda B, Gazquez C, Elosegui-Artola A, Kim DS, De Juan-Pardo E, Demeyer K, Hole K, Larrea E *et al.*: **N-terminal acetylome analyses and functional insights of the N-terminal acetyltransferase NatB**. *Proc Natl Acad Sci U S A* 2012 http://dx.doi.org/10.1073/pnas.1208616109.
Van Damme and colleagues combine quantitative N-terminal enrichment with knockout and knockdown studies in yeast and human for the first in depth characterization of the N-acetyltransferase.

37. Bienvenut WV, Sumpton D, Martinez A, Lilla S, Espagne C, Meinnel T, Giglione C: **Comparative large scale characterization of plant versus mammal proteins reveals similar and idiosyncratic N-α-acetylation features**. *Mol Cell Proteomics* 2012, **11** M111.015131.

38. Chen T, Muratore TL, Schaner Tooley CE, Shabanowitz J, Hunt DF, Macara IG: **N-terminal α-methylation of RCC1 is necessary for stable chromatin association and normal mitosis**. *Nat Cell Biol* 2007, **9**:596-603.

39. Petkowski JJ, Schaner Tooley CE, Anderson LC, Shumilin IA, Balsbaugh JL, Shabanowitz J, Hunt DF, Minor W, Macara IG: **Substrate specificity of mammalian N-terminal α-amino methyltransferase NRMT**. *Biochemistry* 2012 http://dx.doi.org/10.1021/bi300278f.

40. Cynis H, Scheel E, Saido TC, Schilling S, Demuth H-U: **Amyloidogenic processing of amyloid precursor protein: evidence of a pivotal role of glutaminyl cyclase in generation of pyroglutamate-modified amyloid-β**. *Biochemistry* 2008, **47**:7405-7413.

41. Resh MD: **Trafficking and signaling by fatty-acylated and prenylated proteins**. *Nat Chem Biol* 2006, **2**:584-590.

42. Pierre M, Traverso JA, Boisson B, Domenichini S, Bouchez D, Giglione C, Meinnel T: **N-myristoylation regulates the SnRK1 pathway in Arabidopsis**. *Plant Cell* 2007, **19**:2804-2821.

43. Schilling O, Barré O, Huesgen PF, Overall CM: **Proteome-wide analysis of protein carboxy termini: C terminomics**. *Nat Methods* 2010, **7**:508-511.

44. Van Damme P, Staes A, Bronsoms S, Helsens K, Colaert N, Timmerman E, Aviles FX, Vandekerckhove J, Gevaert K: **Complementary positional proteomics for screening substrates of endo- and exoproteases**. *Nat Methods* 2010, **7**:512-515.

45. Wu J, Tolstykh T, Lee J, Boyd K, Stock JB, Broach JR: **Carboxyl methylation of the phosphoprotein phosphatase 2A catalytic subunit promotes its functional association with regulatory subunits in vivo**. *EMBO J* 2000, **19**:5672-5681.

46. • Kornberg TB: **Barcoding Hedgehog for intracellular transport**. *Sci Signal* 2011, **4**:pe44.
Revisiting hedgehog signaling Thomas Kornberg discusses how N-terminal and C-terminal modification regulates intracellular and extracellular movement of proteins.

47. Strahl BD, Allis CD: **The language of covalent histone modifications**. *Nature* 2000, **403**:41-45.

48. Garske AL, Oliver SS, Wagner EK, Musselman CA, LeRoy G, Garcia BA, Kutateladze TG, Denu JM: **Combinatorial profiling of chromatin binding modules reveals multisite discrimination**. *Nat Chem Biol* 2010, **6**:283-290.

49. • Rando OJ: **Combinatorial complexity in chromatin structure and function: revisiting the histone code**. *Curr Opin Genet Dev* 2012, **22**:148-155.
Oliver Rando discusses the ins and outs of combinatorial use of PTMs and the contradicting evidences for the importance of their relevance on histones providing excellent base for the study of combinatorial PTMs in other areas.

50. •• Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, Kellie JF, Li M *et al.*: **Mapping intact protein isoforms in discovery mode using top-down proteomics**. *Nature* 2011, **480**:254-258.
In this excellent work at the cutting edge of top-down proteomics the authors report the first true large-scale top-down analysis identifying >3000 protein species. This work also highlights the prevalence of diversification by posttranslational modifications.

51. O'Brien RJ, Wong PC: **Amyloid precursor protein processing and Alzheimer's disease**. *Annu Rev Neurosci* 2011, **34**:185-204.

52. De Strooper B, Annaert W: **Novel research horizons for presenilins and γ-secretases in cell biology and disease**. *Annu Rev Cell Dev Biol* 2010, **26**:235-260.

53. Jefferson T, Čaušević M, auf dem Keller U, Schilling O, Isbert S, Geyer R, Maier W, Tschickardt S, Jumpertz T, Weggen S *et al.*: **Metalloprotease meprin beta generates nontoxic N-terminal amyloid precursor protein fragments in vivo**. *J Biol Chem* 2011, **286**:27741-27750.